Not Everything Is Greenwashing: Limitations of Automatic Analysis of Sustainability Reports, and a Proposal

Maria Pilar Uribe Silva, Rik van Noord, Malvina Nissim
CLCG, University of Groningen
{m.p.uribe.silva, m.nissim, r.i.k.van.noord}@rug.nl

Abstract

Sustainability reports (SRs) are essential for holding companies accountable, and they are required by law. They also serve as a key communication tool through which companies shape their image and disclose non-financial information. However, the rapid growth of these reports, their lack of standardisation, and the frequent use of strategically ambiguous language make it difficult for stakeholders to evaluate whether sustainability claims are genuine or deceptive. Previous work has focused on extracting misleading climate-related content and identifying greenwashing. We argue that this is not enough, because deception does not only appear in overtly false or misleading green claims, and it often emerges through a variety of subtle linguistic strategies. We therefore propose the development of a framework based on deception theories to examine how deceptive language operates in SRs, and we outline the challenges that should be taken as an invitation for future research.

Keywords: Deception, Deceptive Language, Sustainability Reports, Greenwashing

1. Introduction and Motivation

Addressing the climate and human rights crises has been recognised as a legal obligation by the Inter-American Court of Human Rights[1], not only for states but also for companies. Decisions such as this reinforce the responsibility of corporations to demonstrate transparency and accountability in their environmental and social actions. In this context, Sustainability Reports (SRs)[2] have become central – and mandatory by law – to how companies present their commitments and actions toward environmental and social goals. For companies, these reports are also valuable corporate assets (Dhir, 2005). Through them, they shape reputation, build trust among stakeholders (Tata and Prasad, 2014; Diouf and Boiral, 2017), and provide financial and non-financial information that supports and impacts investment decisions (Abideen and Fuling, 2024). On the other hand, for stakeholders, including capital markets, governments, and civil society, these documents are key tools for assessing whether corporate actions align with stated sustainability commitments (Ellimäki et al., 2023).

Over the past decade, both the number and diversity of SRs have grown considerably. According to a 2024 survey[3], 96% of the world's 250 largest companies now publish such reports. At the same time, the language of these documents has become increasingly complex. Research highlights issues of readability (Smeuninx et al., 2020; Pombinho et al., 2024), lexical consistency (Chalmers and Klingler-Vidra, 2023), and performance measurement (Wang et al., 2018). Despite the efforts of global frameworks such as the Global Reporting Initiative (GRI), the UN Sustainable Development Goals (SDGs), and the Sustainability Accounting Standards Board (SASB), there is still no mandated format or shared terminology, which leaves room for companies to adopt discourse strategies that appear responsible and positive – a well-known practice called *greenwashing* – while avoiding concrete accountability.

Parallel to the growth of SRs, research on misleading language in sustainability communication has also grown, following two main paths. One path has seen the surge of **manual** and corpus-assisted discourse studies, which have provided fine-grained insights into how deception is constructed linguistically: vagueness, exaggeration, and selective disclosure, for instance. These have contributed to revealing the complexity and persuasive tone of corporate communication (Christensen et al., 2013; Elving et al., 2015; Siano et al., 2017; Nwagbara and Belal, 2019; Parlakkaya and Stenka, 2024; Kochkina et al., 2024), and how language is actively used to create social realities, rather than simply reflecting them (Ashcraft et al., 2009).

The other path concerns the development of **automatic** detection methods, based on natural language processing (NLP) techniques, which have

[1] https://www.corteidh.or.cr/docs/opiniones/seriea_32_en.pdf

[2] For simplicity, we will refer to Sustainability Reports (SRs) as a general term throughout this paper. These include Environmental, Social, and Governance (ESG) reports, Corporate Social Responsibility (CSR) reports, Integrated Reports (IR) and other related titles used by organisations to disclose their sustainability-related activities and performance.

[3] https://kpmg.com/dk/en/esg/survey-of-sustainability-reporting-2024.html

Proceedings of the 2nd Workshop on Ecology, Environment, and Natural Language Processing (NLP4Ecology) @ LREC 2026, pages 154–162
May 12, 2026

contributed to scaling up the analysis of corporate disclosures. The majority of them, however, have focused primarily on detecting climate-related claims and greenwashing, which is too broad and unspecific to capture the complex and strategic nature of misleading corporate language. Moreover, they concentrate on the content of the statements rather than on their structure or form, where deception often lies. This limitation is evident in traditional NLP methods, which require clear, observable, and operational linguistic categories, rather than a polymorphic and dynamic term (Calamai et al., 2026; Gorovaia and Makrominas, 2025; Sneiderienė and Legenzova, 2025; Stammbach et al., 2023; Schimanski et al., 2024; Mohammadrezaei et al., 2024). Modern LLM-based approaches present other shortcomings: while they are able to detect explicit environmental claims and inconsistencies at scale, their performance drops when dealing with subtle, context-dependent strategies such as selective disclosure, vague commitments, and aspirational rhetoric (Vel utharambath et al., 2026; Salminen et al., 2025). In addition, these models may align with prevalent misconceptions (Vel utharambath et al., 2026). A statement can be factually correct and still be deceptive. For example, the sentence *"We reduced our carbon emissions by 20% compared to last year"* may be factually correct, but is deceptive if the reduction was due to the temporary closure of a facility, while overall long-term emissions continue to rise.

Contributions We believe the current NLP approaches to sustainability reporting analysis would be much more impactful by taking into account the wealth of research on deceptive language in communication, corporate communication, and linguistics. In this **short position paper**, we propose examining how deceptive language manifests in these types of documents and moving beyond the detection of greenwashing to the detection of broader categories of deceptive linguistic cues in SRs. We highlight the need for a conceptual framework grounded in deception theories and corporate strategies to guide their identification. This shift towards a transdisciplinary approach helps us see how deception operates in SRs, not only what these reports claim but also how they say it.

2. Related work and our position

The growing body of SRs highlights a clear need for more systematic and scalable evaluation. In this section, we discuss the foundational contribution of manual analysis techniques to detect linguistic cues of deception in the development of automated approaches. We then argue why this is insufficient and requires further advances, starting from a better framing of the problem.

Manual analysis Until now, a close reading analysis has played a fundamental role in uncovering corporate discourse strategies. As Siano et al. (2017) demonstrate in their content analysis of the Volkswagen emissions scandal[4], deception in corporate discourse often takes the form of what they call "deceptive manipulation": a new type of irresponsible and deliberate corporate practice. Their work organised different kinds of greenwashing into a taxonomy that they then expanded, bringing two crucial insights: first, that contextual and critical reading helps uncover how companies use complex discursive strategies (Catenaccio et al., 2012; Painter and Martins, 2017) to construct a gap between discourse and action (Elving et al., 2015); and second, that breaking down this broad concept is essential to capturing the complexity of such deceptive practices. Unfortunately, the quantity and persuasive tone of these reports, combined with their qualitative structure and legal complexity, make them particularly difficult and time-consuming to analyse manually, even for professionals who have been working in the sector for a long time. For this reason, developing automatic methods to analyse these kinds of documents effectively and efficiently has become essential.

NLP methods Given the above, several studies have focused on using NLP methods to detect misleading practices, particularly greenwashing. Polignano et al. (2022) developed a system that uses NLP and information extraction methods to identify references to various sustainability topics within reports that comply with Global Reporting Initiative standards (GRI), improving the accessibility of these disclosures. Similarly, Gutierrez-Bustamante and Espinosa-Leal (2022) employed text mining techniques to evaluate the consistency of Nordic companies' sustainability reports with the GRI framework, demonstrating the advantage of models such as Latent Semantic Analysis (LSA) and Global Vectors for Word Representation (GloVe) in this context. Luccioni et al. (2020) presented ClimateQA[5], a customised NLP model designed to identify climate-relevant sections through a question-answer approach, facilitating the extraction of relevant information. Gorovaia and Makrominas (2025) employed text analysis of corporate reports to identify inconsistencies in reporting between companies that had committed environmental violations and those that had not.

More recently, Calamai et al. (2026) reviewed 61 studies addressing one or more aspects of greenwashing detection in text. Their findings show that there are no datasets explicitly labelled for green-

[4] https://www.epa.gov/vw/learn-about-volkswagen-violations

[5] https://www.climateqa.com/

Page	Text excerpt	Deceptive	Greenwashing
17	We are ensuring that we have **a well-balanced, healthy and relevant store portfolio in each market**.	Yes	**No**
31	[...] **improve the resilience** of our business and supply chain, and **empower our customers to be part of our industry's transformation**	Yes	**No**
45	H&M Group is a **customer-focused, creative, value-driven, responsible fashion** and design company	Yes	Yes
53	Good procurement practices and close cooperation with suppliers are essential to ensuring that the company's products are always produced with **consideration for people and the environment.**	Yes	Yes
81	We are committed to **growing our business in a sustainable way** — by decoupling our growth from the use of finite natural resources	Yes	Yes
143	Our vision to **lead the change** means innovating, incubating and investing in **scaling new materials.**	Yes	**No**
143	Providing **convenient ways for our customers to engage** in circular fashion.	Yes	Yes

Table 1: Examples of deceptive statements from the H&M Sustainability Report (2023). Not all of them meet the definition of greenwashing in the European Union's Green Claims Directive Proposal.

washing (also due to the ambiguity of the term), and many current studies fragment global detection into intermediate tasks (e.g., detection of climate-related claims (Stammbach et al., 2023), identification of statements that may indicate greenwashing (Bingler et al., 2024), and tone analysis (Kang and Kim, 2022; Polignano et al., 2022).

Undoubtedly, these studies have helped to extend the analysis beyond what manual approaches can do. But the fundamental problem remains unresolved: inconsistencies and ambiguities in the use of language in SRs remain a challenge in judging whether a public statement reflects a genuine commitment or not. Without a clear conceptual understanding of how deception works in corporate language, it is hard to tell the difference between strategic ambiguity and genuine uncertainty.

LLM-based tools The widespread use and continuous development of Large Language Models (LLMs) have brought them into focus as valuable tools in detecting misleading information in SRs. The ESGReveal system proposed by Zou et al. (2023) enhances LLMs with Retrieval-Augmented Generation (RAG) to retrieve structured ESG information, offering a benchmark for corporate reporting. Bronzini et al. (2024) used LLMs to construct knowledge graphs to analyse the disclosure of ESG in SRs. The work done by Usmanova and Usbeck (2024) proposes a conceptual framework for subdomain-specific pre-training of LLMs to improve the detection of green claims in corporate reports. And recently, Mishra et al. (2024) started offering an open-access and collaborative platform, DocQA[6], for extracting information from technical and unstructured documents, such as SRs, through a conversational assistant that answers questions.

While the potential of this technology is clear, important challenges remain. These models can flag potential cases of greenwashing and even spot certain linguistic cues of deception, but their analyses remain vague, show a tendency toward truth bias (Velutharambath et al., 2026; Markowitz and Hancock, 2023), and often lack adequate explanation. This vagueness does not only stem from the current technical limitations of LLMs, but it is also a product of imprecise language. Model performance is likely negatively affected by the use of existing definitions of greenwashing, which are at the same time overly broad and too limited to be effective in a highly domain-specific task such as SR analysis (Calamai et al., 2026). As a result, automated systems may flag the wrong signals or overlook more subtle forms of deception in SRs.

The problem Despite the valuable contributions of the aforementioned works, the lack of a clear, holistic framework that draws from multiple disciplines to identify deceptive discourse strategies within sustainability narratives at scale undermines the goal of analysing the fast-growing number and size of SR documents. This is where the core of the issue lies: **detecting deception in corporate sustainability disclosures requires more than extracting information to identify greenwashing**. As shown in Table 1, the definition of greenwashing according to the European Union's Green Claims Directive Proposal[7] does not cover deceptive linguistic strategies that appear in sustainability-related publications. Vague or non-specific statements, as well as the omission of relevant information, can function as deceptive practices even when no explicit false claim is made. This shows that focusing only on explicit claims is not enough to capture how deception appears in these texts. Such complexity demands a deeper understanding of how language is used by companies to blur boundaries.

[6] https://ds4sd.github.io/

[7] See § 1.2: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2023:0166:FIN

Our proposal Instead of focusing on greenwashing or the detection of climate-related claims, we propose to concentrate on deceptive corporate strategies. We argue that deceptive language does not always appear in overtly false or misleading green claims; rather, it emerges through a variety of subtle linguistic strategies that distort or obscure meaning. This issue has been extensively explored in the field of deception studies from a linguistic perspective, which provides valuable theoretical and methodological insights for developing a conceptual framework that makes the detection of deception in SRs more precise, operationalisable, and critically informed. By operationalising such a framework into automatic tools, we can move toward a more systematic and accurate analysis of deceptive language in SRs. As an initial step, this can be approached as a sentence and paragraph-level classification task focused on identifying observable deceptive discourse cues.

We draw on deception theory as a conceptual lens, but we do not assume that deceptive intent can be directly recovered from text alone. What can be operationalised are **observable linguistic and discourse phenomena**, such as vagueness, equivocation, obfuscation, and selective disclosure, which may function as deceptive strategies in a corporate context. We therefore distinguish between textual evidence and the stronger interpretive claim that such evidence reflects deceptive intent.

In the next section, we discuss the main theories of deception that underpin our position and explain the opportunity to create a conceptual framework grounded in deception theories, illustrating how deception manifests itself in corporate sustainability communication. In Section 4 we then review the main challenges ahead for operationalising and practically implementing such a framework.

3. Deception as Communicative and Linguistic Strategy

When we talk about deception in language, we talk about *intention*. The intention to hide information, to be ambiguous, and to avoid full cooperation in communication. But intention is not directly observable and cannot be reliably identified from language alone. Even so, research on deception has developed conceptual distinctions and methodological approaches that help scholars define deception, describe its main forms, and analyse how it may be realised in discourse.

We draw on these theories to examine what deception may look like in the context of SRs, how it may operate, and which linguistic and discourse-level cues may function as deceptive strategies in this setting. These include forms such as vagueness, equivocation, selective disclosure, and obfuscation. In this sense, deception theory serves as a conceptual lens, while the analysis itself focuses on observable textual evidence rather than on directly inferring hidden intent.

Theoretical foundation Galasinski (2000) encourages a shift away from the conventional understanding of deception as outright falsehoods. Instead, he conceptualises deceptive communication as a form of linguistic manipulation that operates within the limits of truth and falsehood. From this perspective, deception is not simply about lying but involves more complex and indirect forms of meaning-making. It can therefore be understood as a **spectrum** of possible manifestations rather than a single act, where *intention* plays a central role (DePaulo et al., 1996; Masip et al., 2004; Mahon, 2007, 2015).

Ekman (1996) identifies deception as a communicative act that involves managing information, behaviour, and emotional expression to influence others' perceptions. Masip et al. (2004) further stress that deception is not the result of misunderstanding or error but a deliberate act of fabricating or manipulating factual or emotional information for strategic purposes. Buller and Burgoon (1996) in their Interpersonal Deception Theory (IDT) emphasise that deception is an interactive process in which both deceiver and receiver dynamically adapt to one another, rather than a unidirectional act.

These perspectives are relevant because they frame deception as a strategic, interactional, and goal-oriented act. When applied to the communicative processes of SRs, companies can be viewed as the senders (the deceivers) and stakeholders as the receivers (the target), where the senders deliberately seek to shape perceptions and gain advantage through selective or distorted communication.

In this sense, deception in language can be viewed as a strategic form of message design and impression management (Buller and Burgoon, 1996; Ekman, 1996), and it can take many forms. It may include, for example, exaggerations, subtle lies (DePaulo et al., 1996); falsification, concealment and equivocation (Buller and Burgoon, 1996); misdirection, strategic ambiguity and deflection (Smith et al., 2010); and vagueness and evasiveness (Galasinski, 2000). As Vrij (2008) points out, these many forms of deception are especially difficult to define and even harder to detect, precisely because they are often subtle, context-sensitive, and adaptive. Moreover, we have to add that deceptive communication is determined by cultural norms, language structure and context (Polignano et al., 2022).

Linguistic insights and automatic detection Developing automatic methods to detect and analyse deceptive content is surely challenging, but there is work from the NLP community that shows the potential of this line of research. For example, Levitan et al. (2018), who studied automatic deception detection, analysed a set of linguistic features in interview dialogues. Other research has examined how deception manifests through pragmatic and stylistic cues such as hedging, indirectness, overgeneralisation, and shifts in lexical choice or modality (Hauch et al., 2015, 2016; Zhou et al., 2004). These studies show that deception can be traced through measurable linguistic signals, yet they also highlight the strong context dependence of such patterns.

This issue is discussed in depth in the work of Gillings (2024), who identified markers such as increased certainty or vagueness, and the strategic use of positive sentiment or abstract language in text. However, as Gillings notes, context remains a key limitation in detection. Recent studies further illustrate both the limits and potential of computational approaches: for instance, Velutharambath et al. (2026) demonstrate low cross-domain generalisability of linguistic cues, while Salminen et al. (2025) combine psycholinguistic features with transformer models to detect deceptive reviews.

While linguistic research has provided detailed qualitative insights into how deception operates through language, these findings have not yet been systematically translated into operational frameworks to understand how deceptive language operates in the context of sustainability. As mentioned, automatic systems still rely predominantly on surface-level features (e.g., word frequency, sentiment, syntactic patterns) and fail to integrate deeper discourse or pragmatic mechanisms. We believe that this gap limits their ability to capture context, intent, and communicative strategy, and to model the textual traces through which potentially deceptive discourse is realised.

4. Challenges

We are aware that, even with a strong framework and previous research supporting the conceptualisation of deception in SRs, our proposal still faces significant challenges when looking at the development of automated tools for analysing the deceptive discourse strategies used by companies. Here, we outline a few.

1. Lack of a domain-specific dataset To our knowledge, there is no dataset explicitly annotated for deceptive linguistic strategies in SRs, making it a necessity to create one to apply such a theoretical and conceptual framework. Importantly, our dataset will focus on observable linguistic and discourse cues rather than inferred authorial intent, and should define a clear unit of analysis, for example, at the sentence and paragraph level. The selection, curation, and maintenance of such a dataset is a collaborative and fundamental effort towards the creation of annotated data and, eventually, the training and evaluation of models that identify deceptive language in this context.

Deceptive discourse in corporate communications is often subtle, context-dependent and sector-specific. For this reason, annotation guidelines must clearly distinguish between observable textual evidence and broader interpretative claims about deception. This is important not only for consistency, but also to ensure the task is feasible. Furthermore, future evaluation should assess whether models generalise across different sectors, writing styles, and repeated text, rather than merely learning industry-specific patterns.

2. From PDF to text SRs are often published in highly variable, multimodal formats (graphs, images, infographics, etc.), usually as non-machine-readable PDFs.[8] This makes it difficult to extract structured and comparable information (Peng et al., 2024; Mahadevkar et al., 2024; Gupta et al., 2025). Even if it were possible to extract clean content, the conversion from PDF to plain text poses the risk of losing information and relevant details, such as the position of elements and visual information, where companies also apply deceptive communication strategies.

3. LLM-written SRs LLMs already support many writing processes in the business and corporate sector (Chiarello et al., 2024), and companies will increasingly rely on these systems to produce content and documents such as SRs. This raises an additional challenge for future research. If LLMs begin to shape the language of SRs more strongly, the form and distribution of deceptive discourse cues may also change. Reports may become more standardised, polished, and stylistically homogeneous, which could make deceptive strategies harder to identify. This raises questions such as: how will the use of language, and especially deceptive language, change in these reports in the coming years? And how would this affect LLM-based methods applied in the detection of deceptive language?

[8] https://www.responsibilityreports.com/

4. Human judgment The role of human judgment is a central challenge. Our aim is to identify linguistic cues that may signal deceptive language in SRs and support their automatic analysis. Advances in LLM capabilities offer significant potential for analysing written texts, but these models can still misclassify, especially when dealing with subtle rhetorical and discursive strategies that demand strong contextual and pragmatic understanding (Kaya and Ghosh, 2024; Velutharambath et al., 2026). Even recent methodologies that improve LLM performance and evaluation (Kaya and Ghosh, 2024) do not fully resolve this issue. Human involvement, therefore, remains essential, not only for validating model outputs, but also for interpreting whether observable textual cues justify broader claims about deception. This is particularly important because such claims go beyond the text itself and depend on context, background knowledge, and pragmatic judgment.

5. Conclusion

Research on how to detect deception has always advanced through collaboration and interdisciplinarity. The increasing quantity and complexity of SRs urgently demand these efforts to be renewed through a more scalable and flexible approach. In this short position paper, we respond to that challenge by proposing a broader conceptual framework grounded in deception theories, with the potential to enable more accurate, linguistically informed automatic detection of deceptive language. Rather than focusing only on explicit green claims, we argue for attention to observable discourse cues that may signal deceptive communication in SRs.

Our aim is to understand how deceptive language operates within SRs on a large scale, from linguistic, pragmatic and discourse-analytical perspectives, while keeping the focus on textual evidence rather than inferred intent. Building on previous efforts in automatic deception detection, we invite the NLP community to take a critical discourse perspective on how companies use language in their public-facing statements about sustainability, and how they may shape stakeholders' perceptions through textual strategies such as vague commitments, obfuscation, and selective disclosure, especially when these are difficult to verify through surface-level analysis alone.

6. Bibliographical References

Zain Abideen and Han Fuling. 2024. Sustainability reporting and investor sentiment. A sustainable development approach to Chinese-listed firms. *Journal of Cleaner Production*, 466:142880.

Karen Lee Ashcraft, Timothy R. Kuhn, and François Cooren. 2009. 1 constitutional amendments: "materializing" organizational communication. *Academy of Management Annals*, 3(1):1–64.

Julia Anna Bingler, Mathias Kraus, Markus Leippold, and Nicolas Webersinke. 2024. How cheap talk in climate disclosures relates to climate initiatives, corporate emissions, and reputation risk. *Journal of Banking & Finance*, 164:107191.

Marco Bronzini, Carlo Nicolini, Bruno Lepri, Andrea Passerini, and Jacopo Staiano. 2024. Glitter or gold? Deriving structured insights from sustainability reports via large language models. *EPJ Data Science*, 13.

David B Buller and Judee K Burgoon. 1996. Interpersonal deception theory. *Communication theory*, 6(3):203–242.

Tom Calamai, Oana Balalau, Théo Le Guenedal, and Fabian M. Suchanek. 2026. Detecting greenwashing: A natural language processing literature survey.

Paola Catenaccio et al. 2012. *Understanding CSR discourse: Insights from linguistics and discourse analysis*. Arcipelago edizioni.

Adam Chalmers and Robyn Klingler-Vidra. 2023. Inconsistent language can undermine the credibility of corporate social responsibility reports.

Filippo Chiarello, Vito Giordano, Irene Spada, Simone Barandoni, and Gualtiero Fantoni. 2024. Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation*, 133:103002.

Lars Thøger Christensen, Mette Morsing, and Ole Thyssen. 2013. CSR as aspirational talk. *Organization*, 20(3):372–393.

Bella M DePaulo, Deborah A Kashy, Susan E Kirkendol, Melissa M Wyer, and Jennifer A Epstein. 1996. Lying in everyday life. *Journal of Personality and Social Psychology*, 70(5):979.

Krishna S. Dhir. 2005. The value of language: concept, perspectives, and policies. *Corporate Communications: An International Journal*, 10(4):358–382.

Dominique Diouf and Olivier Boiral. 2017. The quality of sustainability reports and impression management: A stakeholder perspective. *Accounting, Auditing & Accountability Journal*, 30(3):643–667.

Paul Ekman. 1996. Why don't we catch liars? *Social Research*, 63(3):801–817.

Pia Ellimäki, Ruth Aguilera, Nuria Hurtado-Torres, and J. Aragon-Correa. 2023. The link between foreign institutional owners and multinational enterprises' environmental outcomes. *Journal of International Business Studies*, 54:910–927.

Wim Elving, Ursa Golob, Klement Podnar, Anne Ellerup Nielsen, and Christa Thomsen. 2015. Guest editorial: The bad, the ugly and the good: New challenges for CSR communication. *Corporate Communications An International Journal*, 20:118–127.

Dariusz Galasinski. 2000. *The language of deception: A discourse analytical study*. Sage Publications.

Mathew Gillings. 2024. *Corpus Linguistic Approaches to Deception Detection*. Routledge.

Nina Gorovaia and Michalis Makrominas. 2025. Identifying greenwashing in corporate-social responsibility reports using natural-language processing. *European Financial Management*, 31(1):427–462.

Tanay Gupta, Tushar Goel, and Ishan Verma. 2025. Exploring multimodal language models for sustainability disclosure extraction: A comparative study. In *The Sixth Workshop on Insights from Negative Results in NLP*, pages 141–149, Albuquerque, New Mexico. Association for Computational Linguistics.

Marcelo Gutierrez-Bustamante and Leonardo Espinosa-Leal. 2022. Natural language processing methods for scoring sustainability reports—a study of Nordic listed companies. *Sustainability*, 14(15).

Valerie Hauch, Iris Blandón-Gitlin, Jaume Masip, and Siegfried L. Sporer. 2015. Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*, 19(4):307–342. PMID: 25387767.

Valerie Hauch, Siegfried L. Sporer, Stephen W. Michael, and Christian A. Meissner. 2016. Does training improve the detection of deception? a meta-analysis. *Communication Research*, 43(3):283–343.

Hyewon Kang and Jinho Kim. 2022. Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods. *Applied Sciences*, 12(11).

Zeyneb N. Kaya and Souvick Ghosh. 2024. Decoding large-language models: A systematic overview of socio-technical impacts, constraints, and emerging questions.

Nataliia Kochkina, Silvia Macchia, and Michela Floris. 2024. Strategic language use in sustainability reporting: An empirical study. *Sustainability*, 16(23).

Sarah Ita Levitan, Angel Maredia, and Julia Hirschberg. 2018. Linguistic cues to deception and perceived deception in interview dialogues. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1941–1950, New Orleans, Louisiana. Association for Computational Linguistics.

Alexandra Luccioni, Emily Baylor, and Nicolas Duchene. 2020. Analyzing sustainability reports using natural language processing.

Supriya Mahadevkar, Shruti Patil, Ketan Kotecha, Lim Soong, and Tanupriya Choudhury. 2024. Exploring ai-driven approaches for unstructured document analysis and future horizons. *Journal of Big Data*, 11.

James Mahon. 2007. A definition of deceiving. *International Journal of Applied Philosophy*, 21:181–194.

James Mahon. 2015. The definition of lying and deception. *Stanford Encyclopedia of Philosophy*.

David Markowitz and Jeffrey Hancock. 2023. Generative AI are more truth-biased than humans: A replication and extension of core truth-default theory principles.

Jaume Masip, Eugenio Garrido, and Carmen Herrero. 2004. Defining deception. *Anales de Psicología*, 20:147–171.

Lokesh Mishra, Cesar Berrospi, Kasper Dinkla, Diego Antognini, Francesco Fusco, Benedikt Bothur, Maksym Lysak, Nikolaos Livathinos, Ahmed Nassar, Panagiotis Vagenas, Lucas Morin, Christoph Auer, Michele Dolfi, and Peter Staar. 2024. ESG accountability made easy: DocQA at your service. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21):23814–23816.

Mahsa Mohammadrezaei, José Carlos Marques, and Asif Huq. 2024. Use of text mining and natural language processing techniques in analyzing sustainability reports: A systematic literature review and assessment.

Uzoechi Nwagbara and Ataur Belal. 2019. Persuasive language of responsible organisation? a critical discourse analysis of corporate social responsibility (CSR) reports of Nigerian oil companies. *Accounting, Auditing & Accountability Journal*, 32(8):2395–2420.

Christopher Painter and Jorge Martins. 2017. Organisational communication management during the Volkswagen diesel emissions scandal: A hermeneutic study in attribution, crisis management, and information orientation. *Knowledge and Process Management (Print Edition): the journal of corporate transformation*, 2017.

Emre Parlakkaya and Renata Stenka. 2024. *Analysing discourse in corporate sustainability reporting: linguistic insights*, pages 456–473.

Jiahui Peng, Jing Gao, Xin Tong, Jing Guo, Hang Yang, Jianchuan Qi, Ruiqiao Li, Nan Li, and Ming Xu. 2024. Advanced unstructured data processing for ESG reports: A methodology for structured transformation and enhanced analysis.

Marco Polignano, Nicola Bellantuono, Francesco Paolo Lagrasta, Sergio Caputo, Pierpaolo Pontrandolfo, and Giovanni Semeraro. 2022. An NLP approach for the analysis of global reporting initiative indexes from corporate sustainability reports. In *Proceedings of the First Computing Social Responsibility Workshop within the 13th Language Resources and Evaluation Conference*, pages 1–8, Marseille, France. European Language Resources Association.

Miguel Pombinho, Ana Fialho, and Jorge Novas. 2024. Readability of sustainability reports: A bibliometric analysis and systematic literature review. *Sustainability*, 16(1).

Joni Salminen, Mekhail Mustak, Soon-Gyo Jung, Hannu Makkonen, and Jim Jansen. 2025. Decoding deception in the online marketplace: Enhancing fake review detection with psycholinguistics and transformer models. *Journal of Marketing Analytics*, pages 1–18.

Tobias Schimanski, Andrin Reding, Nico Reding, Julia Bingler, Mathias Kraus, and Markus Leippold. 2024. Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication. *Finance Research Letters*, 61:104979.

Alfonso Siano, Agostino Vollero, Francesca Conte, and Sara Amabile. 2017. "More than words": Expanding the taxonomy of greenwashing after the Volkswagen scandal. *Journal of Business Research*, 71:27–37.

Nils Smeuninx, Bernard De Clerck, and Walter Aerts. 2020. Measuring the readability of sustainability reports: A corpus-based analysis through standard formulae and NLP. *International Journal of Business Communication*, 57(1):52–85.

Sandi W Smith, Steven R Wilson, Judee K Burgoon, and Timothy R Levine. 2010. Advances in deception detection. In *New Directions in Interpersonal Communication Research*, pages 201–220. SAGE Publications, Inc.

Agne Sneideriene and Renata Legenzova. 2025. Greenwashing prevention in environmental, social, and governance (ESG) disclosures: A bibliometric analysis. *Research in International Business and Finance*, 74:102720.

Dominik Stammbach, Nicolas Webersinke, Julia Bingler, Mathias Kraus, and Markus Leippold. 2023. Environmental claim detection. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1051–1066, Toronto, Canada. Association for Computational Linguistics.

Jasmine Tata and Sameer Prasad. 2014. CSR communication: An impression management perspective. *Journal of Business Ethics*, 132.

Aida Usmanova and Ricardo Usbeck. 2024. Structuring sustainability reports for environmental standards with LLMs guided by ontology. In *Proceedings of the 1st Workshop on Natural Language Processing Meets Climate Change (ClimateNLP 2024)*, pages 168–177, Bangkok, Thailand. Association for Computational Linguistics.

Aswathy Velutharambath, Kai Sassenberg, and Roman Klinger. 2026. What if deception cannot be detected? A cross-linguistic study on the limits of deception detection from text.

Aldert Vrij. 2008. *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons.

Xuepeng Wang, Kang Liu, and Jun Zhao. 2018. Detecting deceptive review spam via attention-based neural networks. In *Natural Language Processing and Chinese Computing: 6th CCF International Conference, NLPCC 2017, Dalian, China, November 8–12, 2017, Proceedings 6*, pages 866–876. Springer.

Lina Zhou, Judee K. Burgon, Douglas P. Twitchell, Tiantian Qin, and Jay F. Nunamaker Jr. 2004. A comparison of classification methods for predicting deception in computer-mediated communication. *Journal of Management Information Systems*, 20(4):139–166.

Yi Zou, Mengying Shi, Zhongjie Chen, Zhu Deng, Zongxiong Lei, Zihan Zeng, Shiming Yang, Hongxiang Tong, Lei Xiao, and Wenwen Zhou. 2023. ESGReveal: An LLM-based approach for extracting structured data from ESG reports. *ArXiv*, abs/2312.17264.